

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May-June 2025			
(B.Tech.) Program: DS/AIML Scheme: IIB			
Regnlag	Examination: LY- Honours	Semester: VIII	
Course Code: HDSC801	Course Name: Text, Web and Social Media Analytics		
Date of Exam: 28/05/25	Duration: 2.5 Hours		Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	Compare Hidden Markov Model (HMM) with Maximum Entropy Markov (MEMMs) Model		CO1	U
b)	Explain the Content Spamming and Link Spamming technique in detail.		CO3	U
c)	Apply the Unsupervised classification learning technique on one of the examples to describe its steps for document sentiment classification.		CO5	Ap
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Discuss the analysis of sequential and navigational patterns w.r.t. web usage mining.		CO3	U
b)	Explain in detail ' Memory-Based Collaborative Filtering '		CO4	U
c)	Illustrate ' Distance-Based ' clustering algorithms in detail.		CO2	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	Give a detailed explanation of ' Opinion Spam Detection '.		CO5	U
b)	Discuss ' Web Usage Mining ' process in detail.		CO3	U
c)	Explain ' Named Entity Recognition (NER) ' approaches for Information Extraction.		CO1	U
Q.4	Solve any two questions out of three. (10 marks each)	20		
a)	Solve and calculate both Normal & More complex versions of inverted indices by explaining the ' Inverted Index ' for the following documents. d1: Diabetes is a chronic condition. d2: Insulin helps control diabetes patients. d3: Healthy diet and exercise can prevent diabetes.		CO3	Ap

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May-June 2025	
(B.Tech.) Program: DS/AIML Scheme: IIB	Examination: LY- Honours Semester: VIII
Course Code: HDSC801	Course Name: Text, Web and Social Media Analytics
Date of Exam: 28/05/25	Duration: 2.5 Hours
Max. Marks: 60	

		CO4	Ap																
b)	What is Behavior Analytic? with the use of the 'Individual Behavior' approach explain in detail about behavior analytics.																		
c)	<p>I) Explain any one of the Text model with example. [5]</p> <p>II) A sample of 1000 patients from a healthcare database with two binary variables: Diabetes (D): Yes or No Obese (O): Yes or No</p> <p>A breakdown of the data is :</p> <table> <tr> <td></td><td>Obese (Yes)</td><td>Obese (No)</td><td>Total</td></tr> <tr> <td>Diabetes(Yes)</td><td>200</td><td>100</td><td>300</td></tr> <tr> <td>Diabetes(No)</td><td>250</td><td>450</td><td>700</td></tr> <tr> <td>Total</td><td>450</td><td>550</td><td>1000</td></tr> </table> <p>[5]</p> <p>i) Find the probability that a patient has diabetes and is obese. ii) Probability that a patient has diabetes. iii) Probability that a patient is obese. iv) Calculate the percentage (%) of obese patients who have diabetes.</p>		Obese (Yes)	Obese (No)	Total	Diabetes(Yes)	200	100	300	Diabetes(No)	250	450	700	Total	450	550	1000	CO2	Ap
	Obese (Yes)	Obese (No)	Total																
Diabetes(Yes)	200	100	300																
Diabetes(No)	250	450	700																
Total	450	550	1000																
